CLINICAL TRIAL PROTOCOL

Protocol Number: Pending

Date: July 11, 2023

Concordance of Molecular Classification Based on Fine Needle Biopsy (FNB) and Surgical Samples

Sponsor: Qualisure Diagnostics Inc.

Medical Expert for the Trial:

Caitlin Yeo, MD, FRCSC
Title: Clinical Instructor, Endocrine Surgeon, University of Calgary
Address: North Tower Room 1006, 1403 29th St NW, Calgary, AB, T2N 2T9
Phone: 403-944-0966
Fax: 403-944-6008
Email: Caitlin.Yeo@albertahealthservices.ca

Investigators Responsible for Conducting the Trial

Caitlin Yeo, Clinical Instructor, Surgery, University of Calgary; Endocrine Surgeon
Adrian Harvey, Clinical Associate Professor, University of Calgary; Endocrine Surgeon
Janice Pasieka, Professor, University of Calgary; Endocrine Surgeon
Shamir Chandarana, Clinical Associate Professor, University of Calgary; Head & Neck Surgeon
Robert Hart, Professor, University of Calgary; Head & Neck Surgeon
Wayne Matthews, Associate Professor, University of Calgary; Head & Neck Surgeon
Martin Hyrcza, Assistant Professor, University of Calgary; Molecular Pathologist
Moosa Khalil, Clinical Associate Professor, University of Calgary; Pathologist
Anthony Magliocco, Pathologist; CEO of Protean BioDiagnostics
Karen Kopciuk, Adjuvant Associate Professor, University of Calgary; Biostatistician, Alberta Health Services
Tasnima Abedin, Biostatistician, Alberta Health Services

Trial Site: Foothills Medical Centre, 3330 Hospital Dr NW, Calgary, AB

Signature Page

By signing below, the Principal Investigator agrees to adhere to the protocol in the conduct of this study. Any change in the study must be reviewed by a formal protocol amendment procedure and the Principal Investigator will submit all changes, amendments and revisions to the Health Research Ethics Board of Alberta Cancer Committee (HREBA.CC). Any change to the protocol that affects patient selection, safety, or changes in the conduct of the trial will require written approval from HREBA.CC and Health Canada before implementing the change.

The Investigator(s) also agree(s) to conduct the study in accordance with the protocol and all applicable laws and regulations including, but not limited to, the International Conference on Harmonization (ICH) Guideline for Good Clinical Practice (GCP), Division 5 of Health Canada's Food and Drug Regulations, the ethical principles that have their origins in the Declaration of Helsinki, and applicable privacy laws.

The Principal Investigator also thereby agrees that HREBA.CC will approve all patient informed consent forms before the study is initiated. The investigator will obtain informed consent and document this process for all patients enrolled on this study.

_____
Principal Investigator (Print)

_____
Signature

_____
Date (dd/mm/yyyy)

STUDY SYNOPSIS

Title	Concordance of Molecular Risk Stratification Based on Fine Needle Biopsy (FNB) and Surgical Samples
Short Title	Molecular Risk by FNA
Protocol Number	HREBA. CC-22-0007
Phase	N/A
Methodology	Prospective blinded comparison study
Study Duration	May 1, 2022 – May 1, 2028
Study Centres	Foothills Medical Centre
Objective(s)	The primary objective is to determine the concordance between Thyroid GuidePx® molecular classifications acquired from fine needle biopsy (FNB) and matched surgical samples.
Number of participants	130
Diagnosis and Main Inclusion Criteria	Diagnosis: Papillary thyroid cancer Main Inclusion Criteria: A diagnosis of papillary thyroid cancer based on an FNB interpreted as a Bethesda V or VI cytology, candidate for resection.
Study Product	Thyroid GuidePx®
Intervention	None
Statistical Methodology	Descriptive statistics will be reported for all study variables. Concordance of molecular classification between FNB and surgical samples will be evaluated by the Cohen kappa statistic. Differences in progression free survival will be evaluated using the log rank statistic. Performance of Thyroid GuidePx® and of the American Thyroid Association risk stratification index will be evaluated by time-related AUROC, sensitivity and specificity.

BACKGROUND INFORMATION

Thyroid cancer is the 8th most common cancer by prevalence, and incidence has been increasing by >6% per year since 1992. Papillary thyroid cancer (PTC) accounts for most thyroid cancers, and the rising incidence of thyroid cancer can be almost entirely attributed to an increased detection of small PTC's (1). PTC usually has a favorable prognosis with high cure rates. However, approximately 10-15% of PTC's display a more aggressive behavior and are often resistant to conventional adjuvant therapies such as radioactive iodine (2,3). Given the increasing number of PTC cases (and the potential burden on health care costs), accurate prognostication is becoming increasingly important. Accurate estimations of prognosis and risk of recurrence will avoid unnecessarily extensive surgery, investigations and follow-up in individuals with a good prognosis, and will direct more extensive surgery, adjuvant therapies and prolonged follow-ups to those who need it.

The American Thyroid Association (ATA) Risk Stratification System

Treatment decisions are informed by the American Thyroid Association (ATA) Risk Stratification System, which estimates the risk of disease recurrence based on clinical and pathological factors. Although this system has been retrospectively validated by a number of studies, the proportion of variance explained is sub-optimal and ranges from 19%-34% (4, 5). The inability of the ATA system to fully predict recurrence may be related to the failure of this system to adequately integrate the risk associated with a tumor's molecular profile. Only one molecular marker, $BRAF^{V600E}$, is currently incorporated in the ATA Risk Stratification System. Another limitation of the ATA Risk Stratification System is that it can only be fully implemented *after* surgery, when all of the clinical and pathological data required to estimate risk are available. At that time, significant treatment decisions have already been made and the impact of surgery has already been realized by the patient.

Estimating Risk Prior to Treatment: The Current Standard of Care

In 2015, the ATA reported on criteria that could be used to select patients for more conservative surgeries such as thyroid lobectomy. According to the 2015 ATA guidelines, patients with the following *preoperative* criteria are considered low risk: tumor size 1 - 4 cm, clinically node negative, no extrathyroidal spread, no family history of thyroid cancer, and no history of radiation therapy to the neck (6). These patients could be initially treated with a thyroid lobectomy (as opposed to a total thyroidectomy). However, 40 - 60% will require a completion thyroidectomy because the *postoperative* risk stratification (using all clinical and pathological information available after surgery) classified them as a higher risk than initially estimated (7-11). In a recent study, 51% of patients required conversion to a completion thyroidectomy because additional risk factors were recognized during surgery or after final pathology was available (12). This demonstrates the limitations of the clinical preoperative risk estimations, and the need for an improved method of preoperative risk estimation. Therefore, thyroid cancer surgeons and patients would benefit from additional information that would refine the decision to perform lobectomy in appropriate patients while minimizing the need for completion thyroidectomy.

Thyroid GuidePx®: A New Molecular Risk Stratification System

The shortcomings of the ATA Risk Stratification System to reliably determine prognosis has been addressed using a new molecular risk stratification system that comprises the Thyroid GuidePx® gene signature. Using a computer software algorithm, transcripts (mRNA) associated with recurrence and transcripts associated with absence of recurrence were identified in a training set of 335 patients with PTC, then validated in a validation set of 167 patients. Tumors included in the dataset were fresh frozen.

There were three molecular subtypes identified based on the pattern of expression of prognostic genes. These were called Type 1, Type 2 and Type 3. Type 3 PTCs had the most aggressive biology and the

highest rate of recurrence. The was subsequently validated in 124 PTC patients from Korea as well as 167 patients from Edmonton.

The molecular subtypes were considered in the context of tumor stage. Specifically, tumors measuring 1 – 4 cm in the absence of lymph node disease (which comprise patients who would be potential candidates for thyroid lobectomy) were separately analyzed from patients with more advanced disease. Interestingly, Type 3 PTCs had a high recurrence rate regardless of tumor stage. Lobectomy candidates with Type 1 and Type 2 PTCs had a recurrence rate of <3% (Figure 1A, B).

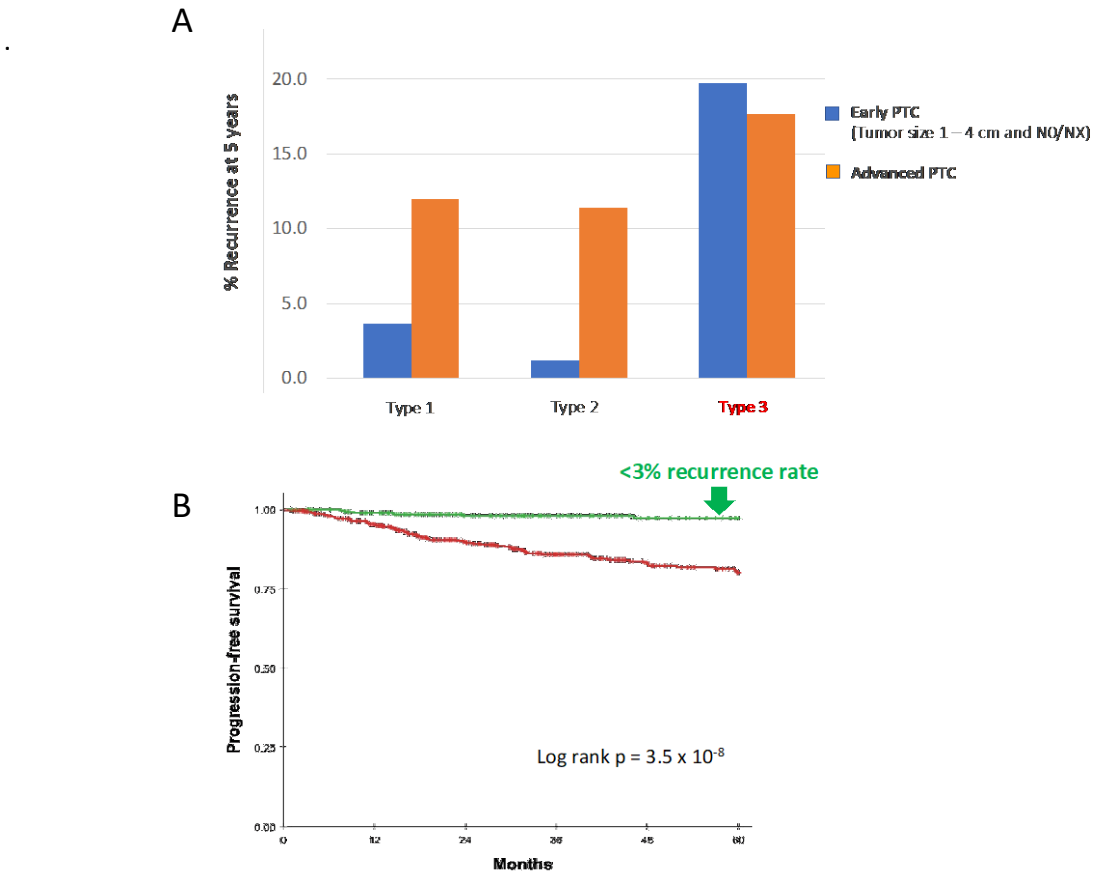


Figure 1. Recurrence outcomes for the three molecular subtypes of PTC. A. Five-year recurrence rates for each molecular subtype as a function of whether tumor is early or advanced. B. Kaplan-Meier curve demonstrating a very low recurrence rate in lobectomy candidates with Type 1 and 2 PTCs. The data are based on a total of 742 patients.

Test Performance

The NGS-based test that identified the molecular subtype of PTC was called Thyroid GuidePx®. The combined molecular and clinical subclassification of PTC using Thyroid GuidePx® was compared against the ATA Risk Stratification System. ATA risk stratification groups were determined by two practicing clinicians for each case. To calculate time-dependent area under the receiver operating characteristic curve (AUROC), a binary classification was applied: low risk versus intermediate/high risk. AUROC was calculated based on recurrence outcomes from the TCGA discovery cohort and the Edmonton cohort. The best risk stratification performance was achieved by the Thyroid GuidePx® test in conjunction with clinical

features (i.e., tumor size and lymph node involvement). Using the smoothROCtime package, the 5-year AUROC was 0.80 for Thyroid GuidePx® when considering tumor size and lymph node status, and 0.51 for the ATA risk stratification system (Figure 2).

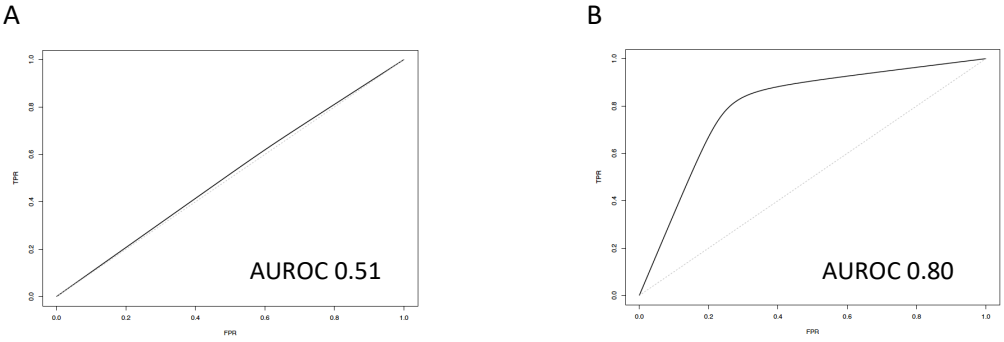


Figure 2. 5-year AUROC for various dichotomous prognostications. The TCGA test cohort (N = 167) and the Edmonton cohort (N = 132) were used. The training set was specifically excluded in this analysis. **A.** ATA risk stratification system (low risk vs. intermediate/high risk). **B**. Low risk defined by Thyroid GuidePx® in conjunction with clinical features (i.e., tumor size and lymph node involvement).

Thyroid GuidePx®: Assay Description

The Thyroid GuidePx® test is an assay based on targeted hybrid-capture enrichment RNA sequencing (RNA-Seq). The test input is RNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue or fine needle biopsy (FNB) papillary thyroid cancer specimens. Library preparation involves specifically amplifying the targeted regions of each of the genes in the Thyroid GuidePx® panel. The panel consists of 82 genes. The resulting sample libraries are transferred to a MiSeq/MiSeqDx reagent cartridge which contains all of the reagents for sequencing. Real-Time Analysis (RTA) software, which is a component of the MiSeq/MiSeqDx sequencer, provides base calls and associated quality scores during the sequencing run. FASTQ files are generated that contain the sequencing and quality information. The FASTQ file is a widely used text-based file format for storing nucleotide sequence for each amplicon and its corresponding quality score. The FASTQ files serve as input for sequence alignment algorithms. Aligned reads are written to files in BAM format which are subsequently converted to read counts and expression values. Gene expression levels are then submitted to the algorithm to classify risk categories.

The software classifies risk based on the pattern of expression of groups of genes within the 82 gene Thyroid GuidePx® panel. Three molecular classes have been defined based on a study with 502 patient samples associating the pattern of gene expression with recurrence free survival at 5 years. A high-risk group enriched with $BRAF^{V600E}$ mutations; and two lower risk molecular subgroups were identified.

Clinical Utility

The more accurate prognostication provided by Thyroid GuidePx® could affect clinicians' treatment decisions. The following are examples of how clinicians may apply the new information derived from Thyroid GuidePx®:

1. Low risk PTC could potentially be treated with a partial thyroidectomy (e.g., lobectomy) or even active surveillance, rather than a total thyroidectomy. This has two major benefits in addition to time and cost savings. First, there is a reduction in the potential need for life-long replacement of

thyroid hormone following a lobectomy. Second, there is virtually no risk of the potentially devastating complications of a total thyroidectomy such as bilateral recurrent laryngeal nerve injury or permanent hypoparathyroidism.

2. Accurately identifying low risk PTC would also increase the appropriateness of adjuvant radioactive iodine (RAI). RAI therapy may result in both transient and long-term side effects such as salivary gland dysfunction, ovarian and testicular dysfunction, and bone marrow suppression. RAI therapy also has a theoretical risk of secondary malignancy: one estimate is that 100mCi ^{131}I induces 56 malignancies per 10,000 patients treated (15). Limiting RAI therapy to those patients who would derive improvement in overall or disease-free survival is also cost effective.
3. Surveillance of PTC can place a strain on patients and healthcare organizations for two reasons: PTC is most commonly diagnosed in middle age, and patients often survive for long periods after their diagnosis. This means that regular examination and investigations for the early detection of recurrence can go on for many years. Follow-up typically involves annual physical examination, serum measurement of thyroid stimulating hormone and thyroglobulin, and periodic neck ultrasound. Low risk patients can have increased intervals between follow-up examinations and investigations, and could potentially be discharged from surveillance.
4. Accurately identifying patients with a high risk of recurrence is also important. The ATA Risk Stratification System is known to be inaccurate at predicting recurrence. In patients classified as "low risk" using the 2015 ATA Risk Stratification System, 16% of patients eventually develop recurrent disease (16). These patients may be undertreated as a result of this risk classification. In patients where a more accurate assignment of high risk is available, more aggressive treatments (i.e. total thyroidectomy and RAI) or enrollment in clinical trials may be considered.

Thyroid GuidePx® by Fine Needle Biopsy

Our studies have so far been done using surgical specimens, and a commercially available assay has been produced in this format. However, a recent study by the Institute of Health Economics has determined that enhanced cost effectiveness of Thyroid GuidePx® would be achieved by performing the test prior to surgery. Implementing a test preoperatively would require that the test performs well using samples derived from a fine needle biopsy (FNB).

Thyroid FNBs are routinely stored in CytoLyt, a methanol based buffered liquid preservative, for up to 7 days prior to transferring cells to a ThinPrep for cytological evaluation. We and others (17–21) have previously determined that quality DNA and RNA can be derived from cells in CytoLyt. Our recent feasibility study consisting of 12 patients with PTC demonstrated that performing the Thyroid GuidePx® assay on FNBs is feasible. So far, the technical success rate has been 100%, and gene level concordance with surgical samples exceeds 0.9. However, reliance on a limited FNB for molecular disease characterization implies that the sample is representative of the entirety of the tumor. Genomic and transcriptomic heterogeneity has been described in primary tumors and metastases (23). Therefore, it will be important to document the concordance between samples acquired by FNB and surgical samples.

TRIAL OBJECTIVES AND PURPOSE

The general purpose of the study is to determine whether the more limited sample from a fine needle biopsy is sufficiently representative of the larger tumor to determine a valid molecular classification using the Thyroid GuidePx® test.

Primary Objective

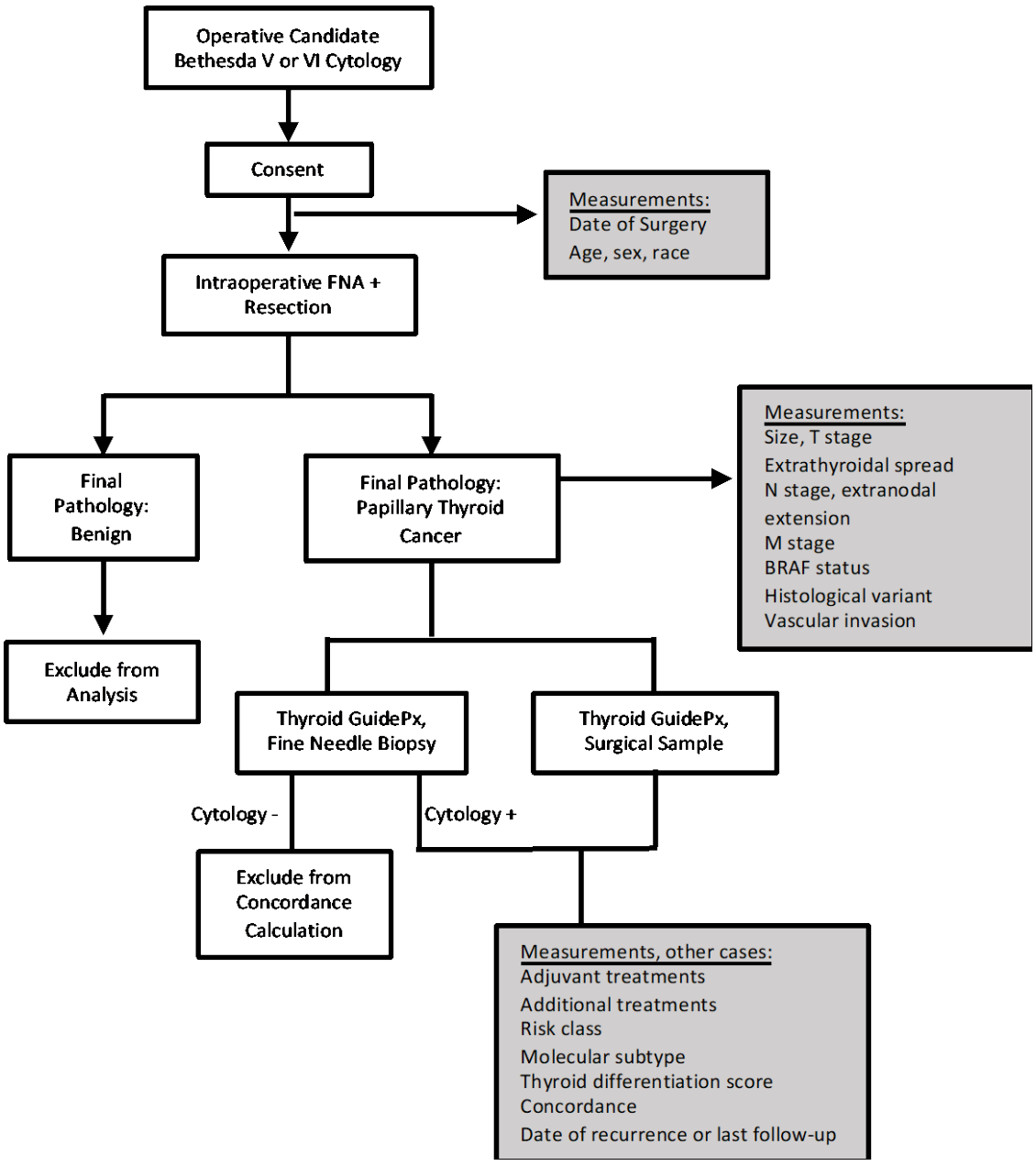
1. To determine the concordance between Thyroid GuidePx® molecular classifications acquired from FNAs and matched surgical samples.

Secondary Objectives

1. To compare the technical success rate in completing a valid Thyroid GuidePx® using FNAs and surgical samples.
2. To compare recurrence outcomes using the ATA risk stratification system vs. patients classified by Thyroid GuidePx® using surgical samples.
3. To compare recurrence outcomes using the ATA risk stratification system vs. patients classified by Thyroid GuidePx® using FNA samples.

TRIAL DESIGN

Trial Schematic



Sample Acquisition and Analysis

Patients will be invited to participate if they have a preoperative tissue diagnosis of PTC (Bethesda VI) or suspicious for PTC (Bethesda V), and they are eligible for partial or total thyroidectomy, as described in the inclusion and exclusion criteria. Following consent, patients will undergo surgery as planned by their surgeon. Demographic, clinical and pathological data will be acquired from all participants.

During surgery, when the thyroid gland and the tumor are exposed, the surgeon will perform an FNB of the dominant tumor (ie: the lesion identified preoperatively), under direct vision. This most closely simulates the *in vivo* conditions of a preoperative FNB yet limits potential technical downfalls of a percutaneous biopsy (eg: inadequate targeting of the lesion). The cellular material derived from the FNB will immediately be put in the transport medium (CytoLyt), then refrigerated. Samples will be sent weekly to the central laboratory for RNA isolation and RNASeq. At the same time, FNB samples will be examined for presence of tumor cells by a trained cytopathologist.

If more than one nodule is present, the FNB nodule will be marked by the surgeon for the pathologist. Under the direction of a collaborating pathologist, a piece of the tumor will be removed and snap frozen, and an adjacent piece will be taken for formalin fixation and paraffin embedding (FFPE). One H & E slide from the FFPE sample with be stained for H & E, and 3 – 6 unstained sections will be prepared, then sent to the central laboratory for microdissection, RNA isolation and RNASeq. This process will follow routine specimen processing protocols and will not interfere with standard methods of pathologic diagnosis.

If patients have signed the optional informed consent form for biobanking, if there is sufficient remaining frozen sample, the sample will be retained for the purpose of re-testing (if needed) or to support additional studies such as whole transcriptome analysis and mutational analysis. Use of these samples for these purposes would require additional ethics approvals.

A feasibility study consisting of 12 patients has already been completed. Technical success rate in acquiring samples and performing the assay was 100%, RNA quantity and quality were excellent in FNBs, and there was excellent gene level concordance in assays performed on FNBs and surgical samples.

RNASeq for Thyroid GuidePx® will be performed at OncoHelix, a clinical laboratory with expertise in precision oncology located at the Cumming School of Medicine, University of Calgary. If testing cannot be done on a timely basis at OncoHelix, samples will be tested at Protean BioDiagnostics, a CAP/CLIA certified lab in Orlando, FL that currently performs the test, commercially. Risk classes will be assigned by software designed by Qualisure Diagnostics using the data generated by RNASeq.

Endpoints

Primary Endpoint:

1. Concordance, molecular risk classification of papillary thyroid cancer FNBs and matched frozen surgical samples

Secondary Endpoints:

1. Technical success rate in completing a valid Thyroid GuidePx® using FNBs
2. Technical success rate in completing a valid Thyroid GuidePx® in FFPE surgical samples.
3. Biochemical recurrence rate at 5 years (thyroglobulin ≥ 1 ng/mL, for patients who had a total thyroidectomy only)
4. Structural recurrence rate at 5 years
5. Test performance as a prognostic test, Thyroid GuidePx®: FNB and surgical samples
    a. Specificity: the proportion of patients without structural recurrence predicted by the test
    b. Sensitivity: the capability of the test to identify individuals whose disease will recur within 5 years (proportion of structural recurrences predicted)
    c. AUROC

d. Positive predictive value
e. Negative predictive value

6. Performance, ATA Risk Stratification
    a. Specificity: the proportion of patients without structural recurrence predicted by the test
    b. Sensitivity: the capability of the test to identify individuals whose disease will recur within 5 years (proportion of structural recurrences predicted)
    c. AUROC
    d. Positive predictive value
    e. Negative predictive value

Blinding and Limitation of Bias

Only samples with sufficient RNA quality (DV200>30) will be included in the analysis. Only FNBs containing tumor cells ("cytology positive") will be included in the concordance calculation. RNASeq will be performed by a technician who is blinded to any clinical details. The technician will specifically be unable to identify matched samples from surgical specimens and FNBs. Molecular classification will be performed in a blinded fashion. Data will be monitored by an independent Data Monitoring Committee. The Data Monitoring Committee, consisting of at least three individuals including a statistician, will monitor data analysis.

Stopping Rules

In our surgical samples, inter-batch and intra-batch repeats consistently had a gene level concordance of >0.9. Gene level concordance between cytology positive FNB and frozen surgical samples (the "gold standard") will be monitored for each batch of 24 patients. If R<0.9, we will pause to re-evaluate the analytical workflow and instrumentation. If concordance is persistently below 0.9 after 72 patients, we will stop evaluating FNBs, but we will continue to test surgical samples. Since this is not a high-risk study, we do not intend to terminate the trial prematurely. Even if there is low concordance between FNB and surgical samples, the data derived from studying outcomes in relation to molecular classification (in surgical samples) will be useful.

SELECTION AND WITHDRAWAL OF SUBJECTS

Inclusion Criteria

1. Age $\geq$ 18 years
2. A diagnosis of papillary thyroid cancer based on a fine needle biopsy (FNB) interpreted as a Bethesda V or VI cytology
3. Tumor size > 1 cm in maximal diameter
4. The patient is an operative candidate
5. The patient has provided consent

Exclusion Criteria

1. Family history of thyroid cancer
2. History of radiation to the neck
3. Unable or unwilling to have a fine needle biopsy
4. Unwilling to undergo thyroidectomy
5. Final pathology does not demonstrate papillary thyroid cancer
6. Cases where there is no clear dominant nodule
7. Cases where there are multiple nodules that preclude sampling of a defined nodule

Subject Withdrawal Criteria

Subjects can discontinue participation at any time by notifying one of the investigators or the research staff. Data acquired prior to withdrawal will be used for analysis, but no additional data will be collected after withdrawal. At the request of the subject, samples can be withdrawn prior to testing. Subjects will be replaced if samples have not been tested by RNASeq (Thyroid GuidePx®).

TREATMENT OF SUBJECTS

The molecular classification and recurrence risk determined by Thyroid GuidePx® will be disclosed to the physician and to the patient. Recurrence risk is based on cumulative validation data including cases from Alberta and Australia. Only the results of the frozen surgical sample (the "gold standard") will be disclosed. As is the usual standard, treatment decisions will be made by the physician in collaboration with the patient, based on a synthesis of all available clinical information. This clinical trial does not define a treatment pathway. Follow up frequency and duration will not be prescribed by protocol and will be determined by the treating physician. This is a non-interventional study, and so compliance will not be monitored.

ASSESSMENT OF EFFICACY

A secondary outcome is test performance to stratify recurrence risk. Test performance will be compared in samples acquired by FNB and surgical samples, then compared to the current standard of care (the ATA risk stratification system). Time to recurrence will be measured from the time of complete initial treatment. In patients with microscopic disease remaining after surgery, time to recurrence will be measured from the completion of the therapeutic dose of radioactive iodine. Recurrence events will be defined based on the 2016 American Thyroid Association guidelines.

The following criteria will signify a disease-free status after therapy: (i) no clinical evidence of tumor; (ii) no evidence of tumor on radioactive iodine imaging and/or cervical ultrasound; (iii) for post-total thyroidectomy, unstimulated serum thyroglobulin (Tg) <0.2 ng/ml; for post-lobectomy, unstimulated Tg<30. An indeterminate response to therapy will be defined as: (i) no clinical evidence of tumor; (ii) no evidence of tumor on radioactive iodine imaging and/or cervical ultrasound; (iii) for post-total thyroidectomy, unstimulated Tg 0.2 – 1; for post-lobectomy, Tg>30. Following a total thyroidectomy, a biochemical incomplete response is defined as Tg>1 in the absence of structural disease. Tg values are only reliable in the absence of interfering antibodies (TgAb) (6).

Disease will be defined as a true recurrence if a patient previously had an undetectable Tg in the absence of TgAb and a negative neck ultrasound examination within 1 year after the previous operation. Persistent disease will be defined as Tg greater than the above thresholds, an abnormal neck ultrasound, persistently increased levels of TgAb, or presence of distant metastases.

Patients with persistent disease and/or metastatic disease (non-curative resection) will not be included in calculation of recurrence-free survival.

ASSESSMENT OF SAFETY

The only departure from standard clinical care is the intraoperative procurement of sample by FNB. However, in comparison to the planned surgical procedure, this risk is small. The main risk is a small risk of clinically significant bleeding, estimated at <1%. As there will not be any intervention associated with this study, there is no plan for monitoring adverse events.

STATISTICS

Optimization of RNASeq Conditions for Samples in CytoLyt

RNA will be isolated from FNBs in CytoLyt as per SOP. RNA quantity will be assessed by NanoDrop and and quality will be quantified by DV_{200}. RNASeq will be performed in CytoLyt samples and matched frozen samples at varying read depths. In paired CytoLyt and FFPE tissues, assay metrics will be evaluated as a function of input tissue storage method, RNA quality (DV_{200}) and storage time. Gene expression will be compared in paired CytoLyt and frozen tissues, then correlated between conditions. The most important outcome, molecular class assignment, will be compared in paired samples.

The following performance metrics will be evaluated for each experiment, based on the College of American Pathologists (CAP) checklist of requirements specific to next generation sequencing:

Assay Metrics

a) RNASeq Metrics:

(i) Read counts (total reads, unique reads, duplicate reads, rRNA reads, strand specificity);

(ii) Coverage (number of reads that cover a given genomic position): mean coverage, mean coefficient of variation, 5'/3' coverage, gaps in coverage, GC bias

(iii) Expression correlation (measured expression levels vs. known levels of ERCC internal standards)

b) Individual Gene Detection:

(i) Analytical sensitivity: fraction of cases in which each gene in the ERCC internal standard is correctly identified.

(ii) Analytical specificity: ability to identify the absence of the genes comprising the ERCC internal standard

(iii) Accuracy: sensitivity + specificity

Comparison of Molecular Risk Stratification (Thyroid GuidePx®) Determined in FNB Specimens and in Surgical Specimens

Patients with a final pathology that does not demonstrate papillary thyroid cancer will be excluded from concordance analysis. Similarly, FNB samples that are cytology negative will be excluded from concordance analysis. The molecular class assignment will be compared between samples obtained by FNB and matched frozen surgical specimens. Surgical samples will be considered the gold standard. The kappa statistic will be used to determine the significance of concordance between the two sample types.

Risk strata will be assigned in a blinded fashion using Thyroid GuidePx® and the ATA Risk Stratification System. Structural and biochemical recurrence rates at 5 years will be compared for each risk class. A binary test result, conditional on disease status (structural recurrence), will be used to assess test accuracies. Structural recurrence represents the gold standard. A 2-sided McNemar test for two dependent proportions will be used to evaluate the sensitivity and specificity of each method of risk stratification, with Holm's adjustment for multiplicity for assessing each accuracy measure separately.

Performance of Thyroid GuidePx®

From the studies described above, the performance of Thyroid GuidePx® will be assessed separately in FNB samples and surgical samples. The performance characteristics will be assessed as follows:

1. Time-dependent AUROC (area under the receiver operator curve) (24). For the purpose of the calculations, the threshold level for a positive test is a designation of intermediate or high risk (since it is anticipated that an intermediate risk assignment would result in significantly different clinical decisions).
2. Sensitivity: calculated as cumulative sensitivity at 5 years, as described by Kamarudin and Kolamunnage-Dona (24)
3. Specificity: calculated as dynamic specificity, as described by Kamarudin and Kolamunnage-Dona (24)

Sample Size Calculation

We will compare the molecular risk assignment from surgical specimens and matched FNBs from 109 patients. A minimum level of concordance that is clinically acceptable is 80%. Sample size calculations were based on the Lin's Kappa statistic for agreement between the two measures for three molecular subgroups (25). A two-sided hypothesis was applied with a significance level of 0.05, and power was fixed at 90%. Molecular subgroup sizes were fixed at 23% (Type 1), 49% (Type 2), and 28% (Type 3). If the null hypothesis is that the level of agreement between FNB and surgical samples is 0.6, then we would be able to detect a level of agreement of 0.8 or higher with 109 patients. A concordance of 0.90 would be detectable with 40 patients assuming the same null level of agreement.

About 25% of patients with Bethesda V cytology and about 5% of patients with a Bethesda VI cytology will have benign disease. In addition, in approximately 5% of cases, RNASeq quality is insufficient in FFPE samples. Based on this, we plan a drop-out rate of 20%. Therefore, the trial is designed to include 130 patients.

DIRECT ACCESS TO SOURCE DATA/DOCUMENTS

The clinical trial site will permit trial-related monitoring, audits, REB review, and regulatory inspection(s), providing direct access to source data/documents.

Clinical trial audits are performed to provide assurance that the rights, safety and wellbeing of patients are properly protected, to assess compliance with the protocol, processes and agreements, ICH Good Clinical Practice standards and applicable regulatory requirements, and to assess the quality of data.

Any domestic regulatory agency may come at any time to audit or inspect the study site. This audit consists of interviews with the principal investigator and study team, review of documentation and practices, review of facilities, equipment and source data verification.

The investigator will grant to any domestic regulatory agency direct access to paper and/or electronic documentation pertaining to the clinical study (e.g. CRFs, source documents such as hospital patient charts and investigator study files). All site facilities related to the study conduct could be visited during an audit (e.g. laboratory, outpatient department). The investigator agrees to co-operate and provide assistance at reasonable times and places with respect to any auditing activity.

QUALITY CONTROL AND QUALITY ASSURANCE

The Principal Investigator will ensure that all study members are following the protocol and complying with regulatory and Good Clinical Practice standards. All clinical trial sites will permit trial-related monitoring, audits, REB review, and regulatory inspection(s), providing direct access to source data/documents.

The Data Monitoring Committee (DMC) will meet during the conduct of the trial as follows: 1. Upon trial initiation to finalize the DMC Charter; 2. After 24 participants have been enrolled; 3. After 72 participants have been enrolled; and 4. after completion of enrollment.

ETHICS

The responsible investigator will ensure that this study is conducted in agreement with the Declaration of Helsinki. This study will be conducted per Canadian and international standards of Good Clinical Practice for all studies including the ICH Harmonized Tripartite Guideline on Good Clinical Practice. Applicable government regulations, NNHPD clinical trial guidelines and Alberta Health Services research policies and procedures will also be followed.

This protocol and any amendments will be submitted to the HREBA-CC for formal approval to conduct the study. The decision of the HREBA-CC concerning the conduct of the study will be made in writing to the investigator.

Disclosure of molecular risk classification from surgical samples (the "gold standard") is justifiable, since there are substantial outcomes data from surgical samples. Moreover, when this trial is launched, the test will be commercially available as a Laboratory Developed Test (LDT).

DATA HANDLING AND RECORD KEEPING

Once the study is completed, all data for this study will be retained for 15 years, stored as per the protocol at the Calgary Centre for Clinical Research.

FINANCING AND INSURANCE

The trial will be funded by Qualisure Diagnostics, Inc., supplemented by peer-reviewed grants. Insurance will be provided by the University of Calgary and Alberta Health Services, as well as Qualisure Diagnostics.

PUBLICATION POLICY

The findings from this clinical study will be published in a peer-reviewed journal with all named investigators as co-authors. All funding sources will be acknowledged.

REFERENCES

1. Davies L, Morris LG, Haymart M, Chen AY, Goldenberg D, Morris J, et al. American Association of Clinical Endocrinologists and American College of Endocrinology Disease State Clinical Review: The Increasing Incidence of Thyroid Cancer. Endocr Pract. 2015;21(6):686-96.

2. Sciuto R, Romano L, Rea S, Marandino F, Sperduti I, Maini CL. Natural history and clinical outcome of differentiated thyroid carcinoma: a retrospective analysis of 1503 patients treated at a single institution. Ann Oncol. 2009;20(10):1728-35.

3. Pacini F, Castagna MG. Approach to and treatment of differentiated thyroid carcinoma. Med Clin North Am. 2012;96(2):369-83.

4. Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. Thyroid. 2010;20(12):1341-9.

5. Castagna MG, Maino F, Cipri C, Belardini V, Theodoropoulou A, Cevenini G, et al. Delayed risk stratification, to include the response to initial treatment (surgery and radioiodine ablation), has better outcome predictivity in differentiated thyroid cancer patients. Eur J Endocrinol. 2011;165(3):441-6.

6. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid. 2016;26(1):1-133.

7. Cheng SP, Chien MN, Wang TY, Lee JJ, Lee CC, Liu CL. Reconsideration of tumor size threshold for total thyroidectomy in differentiated thyroid cancer. Surgery. 2018;164(3):504-10.

8. Dhir M, McCoy KL, Ohori NP, Adkisson CD, LeBeau SO, Carty SE, et al. Correct extent of thyroidectomy is poorly predicted preoperatively by the guidelines of the American Thyroid Association for low and intermediate risk thyroid cancers. Surgery. 2018;163(1):81-7.

9. Lang BH, Shek TW, Wan KY. The significance of unrecognized histological high-risk features on response to therapy in papillary thyroid carcinoma measuring 1-4 cm: implications for completion thyroidectomy following lobectomy. Clin Endocrinol (Oxf). 2017;86(2):236-42.

10. Kluijfhout WP, Pasternak JD, Lim J, Kwon JS, Vriens MR, Clark OH, et al. Frequency of High-Risk Characteristics Requiring Total Thyroidectomy for 1-4 cm Well-Differentiated Thyroid Cancer. Thyroid. 2016;26(6):820-4.

11. Murthy SP, Balasubramanian D, Subramaniam N, Nair G, Babu MJC, Rathod PV, et al. Prevalence of adverse pathological features in 1 to 4 cm low-risk differentiated thyroid carcinoma. Head Neck. 2018;40(6):1214-8.

12. Craig SJ, Bysice AM, Nakoneshny SC, Pasieka JL, Chandarana SP. The Identification of Intraoperative Risk Factors Can Reduce, but Not Exclude, the Need for Completion Thyroidectomy in Low-Risk Papillary Thyroid Cancer Patients. Thyroid. 2020.

13. The Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014;159(3):676-90.

14. Craig S, Stretch, C, Farshidfar F, Sheka D, Alabi N, kopciuk K, Harvey A, Park YJ, Bathe O. Molecular Subgroups Identified by Machine Learning Enable Accurate Prognostication of Papillary Thyroid Cancer. Presented at the 90th Annual Meeting of the American Thyroid Association (Virtual Meeting; 2021). https://qualisuredx.com/wp-content/uploads/2021/12/2021-10_ATA_Conference_2021_SCraig.pdf

15. Rubino C, de Vathaire F, Dottorini ME, Hall P, Schvartz C, Couette JE, et al. Second primary malignancies in thyroid cancer patients. Br J Cancer. 2003;89(9):1638-44.

16. Pitoia F, Jerkovich F, Urciuoli C, Schmidt A, Abelleira E, Bueno F, et al. Implementing the Modified 2009 American Thyroid Association Risk Stratification System in Thyroid Cancer Patients with Low and Intermediate Risk of Recurrence. Thyroid. 2015;25(11):1235-42.

17. Zhou H, Mody DR, Smith D, Lloyd MB, Kemppainen J, Houghton J, et al. FNA needle rinses preserved in Cytolyt are acceptable specimen type for mutation testing of thyroid nodules. J Am Soc Cytopathol. 2015;4(3):128-35.

18. Fuller MY, Mody D, Hull A, Pepper K, Hendrickson H, Olsen R. Next-Generation Sequencing Identifies Gene Mutations That Are Predictive of Malignancy in Residual Needle Rinses Collected from Fine-Needle Aspirations of Thyroid Nodules. Arch Pathol Lab Med. 2018;142(2):178-83.

19. Chen T, Gilfix BM, Rivera J, Sadeghi N, Richardson K, Hier MP, et al. The Role of the ThyroSeq v3 Molecular Test in the Surgical Management of Thyroid Nodules in the Canadian Public Health Care Setting. Thyroid. 2020;30(9):1280-7.

20. Eszlinger M, Bohme K, Ullmann M, Gorke F, Siebolts U, Neumann A, et al. Evaluation of a Two-Year Routine Application of Molecular Testing of Thyroid Fine-Needle Aspirations Using a Seven-Gene Panel in a Primary Referral Setting in Germany. Thyroid. 2017;27(3):402-11.

21. Titov S, Demenkov PS, Lukyanov SA, Sergiyko SV, Katanyan GA, Veryaskina YA, et al. Preoperative detection of malignancy in fine-needle aspiration cytology (FNAC) smears with indeterminate cytology (Bethesda III, IV) by a combined molecular classifier. J Clin Pathol. 2020;73(11):722-7.

22. Gonzalez HE, Martinez JR, Vargas-Salas S, et al. A 10-gene classifier for indeterminant thyroid nodules: development and multicenter accuracy study.Thyroid 2017; 27(8).

23. Le Pennec S, Konopka T, Gacquer D, Fimereli D, Tarabichi M, Tomas G, et al. Intratumor heterogeneity and clonal evolution in an aggressive papillary thyroid cancer and matched metastases. Endocr Relat Cancer. 2015;22(2):205-16.

24. Kamarudin AN, Cox T, Kolamunnage-Dona R: Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Med Res Methodol 2017, 17(1):53.

25. Flack V, Afifi A, Lachenbruch P, Schouten H. Sample size determinations for the two rater kappa statistic. Psychometrika. 1988;53(3):321-5.